Bubble coalescence in breathing DNA: Two vicious walkers in opposite potentials

Tomáš Novotný[1 2], Jonas Nyvold Pedersen[3], Tobias Ambjörnsson[4 5], Mikael Sonne Hansen[6], and Ralf Metzler[4 7]

[1] *Nano-Science Center, University of Copenhagen - Universitetsparken 5, 2100 Copenhagen, Denmark*
[2] *Dept. of Condensed Matter Physics, Faculty of Mathematics and Physics, Charles University - Ke Karlovu 5, 121 16 Prague, Czech Republic*
[3] *Mathematical Physics, Lund University - Box 118, 22100 Lund, Sweden*
[4] *NORDITA - Blegdamsvej 17, 2100 Copenhagen, Denmark*
[5] *Dept. of Chemistry, Massachusetts Institute of Technology - 77 Massachusetts Ave, Cambridge, MA 02139, USA*
[6] *Dept. of Mathematics, Technical University of Denmark - Bldg. 303S, Matematiktorvet, 2800 Kgs. Lyngby, Denmark*
[7] *Physics Dept., University of Ottawa - 150 Louis Pasteur, Ottawa, ON, K1N 6N5, Canada*

PACS. 87.14.Gg – DNA, RNA.
PACS. 02.50.Ey – Stochastic processes.
PACS. 82.37.-j – Single molecule kinetics.

Abstract. – We investigate the coalescence of two DNA-bubbles initially located at weak segments and separated by a more stable barrier region in a designed construct of double-stranded DNA. The characteristic time for bubble coalescence and the corresponding distribution are derived, as well as the distribution of coalescence positions along the barrier. Below the melting temperature, we find a Kramers-type barrier crossing behaviour, while at high temperatures, the bubble corners perform drift-diffusion towards coalescence. The results are obtained by mapping the bubble dynamics on the problem of two vicious walkers in opposite potentials.

Introduction. – The Watson-Crick double helix is the thermodynamically stable state of double-stranded DNA in a wide range of temperatures and salt conditions [1]. This stability is effected by hydrogen bonds between the bases in individual base-pairs (bps), and the stronger stacking interactions between nearest neighbour pairs of bps. Driven by thermal fluctuations double-stranded DNA can break apart, to form denaturation bubbles of flexible single-stranded DNA [2]. Although rare, bp-opening events expose active groups of DNA bases, that are otherwise buried within the double helix. They are crucial for the interaction with proteins and chemicals, and therefore for the biological function of DNA. Bubble kinetics has been probed in NMR studies [3], and the growth and reannealing dynamics of individual bubbles has been measured in real time in single DNA fluorescence correlation setups [4].

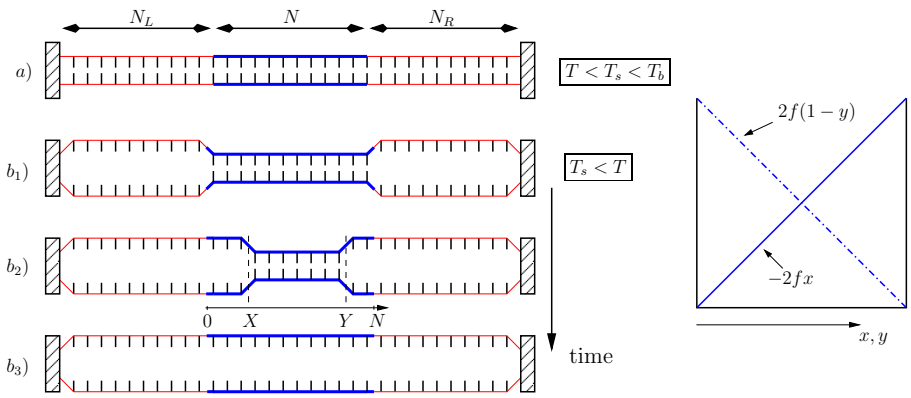

Fig. 1 – **Left:** Schematic of the bubble coalescence setup in a designed DNA construct. It is clamped at both ends and consists of two outer soft zones (thin red lines) of lengths N_L, N_R bps with melting temperature T_s and a stronger N-bps-long barrier zone (thick blue lines) with $T_b > T_s$. a) All bps closed ($T < T_s < T_b$). b) Soft zones open by raising the temperature above T_s. b_1–b_3) Successive opening of the barrier driven mainly by fluctuations ($T < T_b$) or drift ($T > T_b$) until coalescence. The discrete coordinates $X, Y = 0, \ldots, N$ are defined as the positions of the *interfaces* between the closed and broken bps. **Right:** Plot of the linear potentials experienced by the respective bubble interfaces in the case $T < T_b$ ($f < 0$) in terms of the dimensionless quantities x, y, f (see text).

The delicate sensitivity of bubble dynamics and, therefore, bubble nucleation to the local DNA bp-sequence [5,6] suggests a new method to use single molecule tools to obtain independently DNA stability parameters, as sketched in figure 1. A short stretch of DNA, clamped at both ends, is designed such that two soft zones consisting of weaker AT-bps are separated by a more stable barrier region rich in GC bps. For simplicity, we assume that both soft zones and barrier are homopolymers with a bp-dissociation free energy ε' and ε, respectively, and, in accordance with the experimental findings of reference [4], we neglect secondary structure formation in the barrier zone. At temperatures higher than the melting temperature T_s of the soft zones but still lower than the melting temperature T_b of the barrier region, thermal fluctuations will gradually dissociate the barrier, until the two bubbles coalesce. We also study the case when the system is prepared as above and then T suddenly increased such that $T > T_b > T_s$ so that the system is driven towards coalescence. In both cases the two boundaries between bubbles and barrier perform a (biased) random walk in opposite free energy potentials. The quantities of interest are the bubble coalescence time and position. It turns out that this problem can be mapped on a previously unsolved case of two vicious walkers in opposite potentials in one dimension and, therefore, is of interest by itself.

Formulation of the problem. – The statistical weight of the construct before coalescence $\mathscr{Z}_{X,Y} = \left(\xi e^{N_L\beta\varepsilon'}\right) e^{(X-Y+N)\beta\varepsilon} \left(\xi e^{N_R\beta\varepsilon'}\right)$ at $T_b > T > T_s$ involves a cooperativity (ring) factor $\xi \approx 10^{-5}$ for each bubble, and a Boltzmann factor for each broken bp with free energies $\varepsilon' > 0$ and $\varepsilon < 0$, compare reference [6].([1]) Upon coalescence, the boundary free energy corresponding to one factor ξ is released, $\mathscr{Z}_{\text{coal}} = \xi e^{(N_L+N_R)\beta\varepsilon'+N\beta\varepsilon}$, stabilizing the system against immediate transition back to a two-bubble state. It should therefore be possible to experimentally detect the coalescence. In our theoretical approach, this corresponds to an absorbing boundary for the random walking interfaces at $X = Y$.

([1])We denote a stable state by a negative free energy.

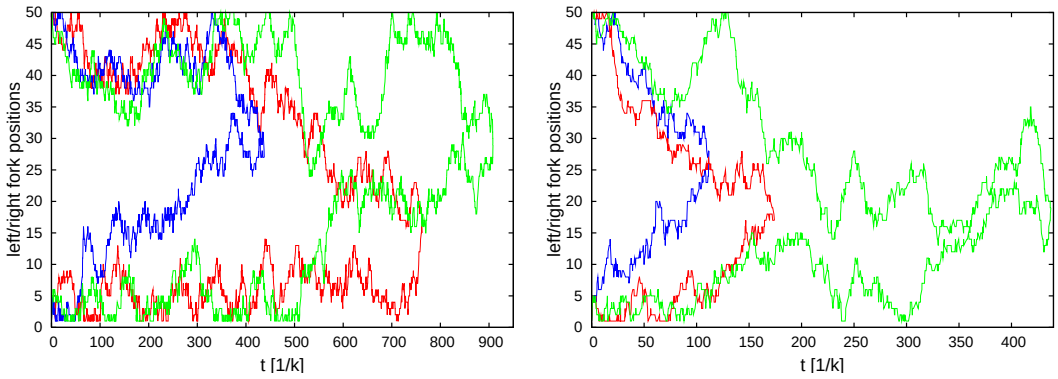

Fig. 2 – Stochastic trajectories of the random walk of the two fork positions encompassing the double-helical bridge of $N = 50$ bps between the vicinal bubbles for $T_s < T < T_b$ with $u = \exp(\beta\varepsilon) = 0.98$ (left) and $T > T_b > T_s$, $u = 1.10$ (right). Note that below the melting temperature T_b of the barrier zone the trajectories have the tendency to move toward the reflecting boundaries at the corners of the soft zone and the coalescence takes longer time due to the presence of the potential barrier. On the other hand, above T_b the trajectories show a tendency to move fast toward the centre of the barrier zone, as expected from the funnelling nature of the positive force directed toward the middle.

The dynamics of this, intrinsically discrete, system can be modelled by methods developed previously, namely, the stochastic Gillespie scheme [7], which allows one to obtain single bubble trajectories and thereby the access to the noise relevant in single molecule experiments, and the master equation approach [6, 8, 9]. As detailed in reference [10], the explicit sequence of bps is included into these approaches. Typical examples of individual trajectories resulting from the Gillespie scheme are displayed in figure 2, where traces of the two interfaces ("forks") cornering the barrier region are shown. Bubble coalescence terminates each pair of trajectories.

Both the Gillespie scheme and the master equation are based on a specific choice for the rate constants, $\mathsf{t}^{(+)-}(Q)$, of (un)zipping the bp at the position $Q = X, Y$ of one of the zipping forks where double-stranded DNA branches into two strands of single-stranded DNA. As discussed in detail in references [6, 8–10], we define

$$\mathsf{t}^{-} = k/2;\ \mathsf{t}^{+}(Q) = ku(Q)s(m)/2,\ s(m) = \{(m+1)/(m+2)\}^{c}, \tag{1}$$

together with the appropriate boundary conditions. Here, t^{-} is the position-independent zipping rate $k/2$ for a bp at the zipping fork, while the bp-unzipping at location Q involves the Boltzmann factor $u(Q) = \exp(\beta\varepsilon(Q))$ for disrupting the bp at Q, as well as the factor $s(m)$ that stems from the entropy loss of forming a closed polymer ring. The s-factor depends on the size m of the bubble cornered by the respective bp, and $c \approx 1.76$ is the critical exponent. In our illustrations and analytic approach we use the reflecting boundary conditions at the edges of the barrier regions, i.e., we assume $\beta\varepsilon' \gg 1$ so that the soft zones are always open.

The rates $\mathsf{t}^{\pm}$ define the transfer matrix $\mathbb{W}$, that governs the random walk of the zipping forks at either end of the barrier region. The coalescence dynamics is then quantified by the probability distribution $P_D(X, Y, t)$ to find the left and right zipping forks at positions X and Y, as controlled by the master equation $\partial P_D(X, Y, t)/\partial t = \mathbb{W}P_D(X, Y, t)$. Solution of this equation by either method, as detailed elsewhere [6–8, 10], yields the quantities of interest such as the mean coalescence time or distribution of the coalescence position.

Continuous description and semi-analytic solution. – Both the master equation and Gillespie approaches can be used for arbitrary bp sequences and their usage is straightforward, however, they both have their limitations. The Gillespie algorithm performs badly in the case of a relatively strong barrier ($\beta\varepsilon \ll -1$) when the coalescence time grows exponentially. This demands exponentially increasing simulation time. The master equation is becoming numerically prohibitive with increasing size of the barrier. On a regular PC the length of about $N \approx 100$ bps is close to the numerical limit of our master equation approach.

For the designed construct we can find an efficient semi-analytical solution for its dynamics. Due to the large length of the soft zones $N_{L,R} \gg 1$ we can well approximate the bubble entropic factors $s(m) = \{(m+1)/(m+2)\}^c$ by 1. Furthermore, for the barrier length $N \gtrsim 50$ we can resort to a continuous description of the fork positions. The discrete master equation can then be rewritten as a bivariate Fokker-Planck equation [11, 12] for the random walk of the two bubble-barrier boundaries characterized by the probability density function (PDF) $P(x, y, t)$ of finding the boundaries at coordinates $x = X/N$ and $y = Y/N$ at a given rescaled time t

$$\frac{\partial}{\partial t}P(x,y,t) = \left(\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right] - 2f\frac{\partial}{\partial x} + 2f\frac{\partial}{\partial y}\right)P(x,y,t), \tag{2}$$

with the dimensionless force $f = N(u-1)/(1+u)$ and time rescaled by $k(1+u)/2N^2$. Equation (2) is completed by the initial condition $P(x, y, 0) = \delta(x - x_0)\delta(y - y_0)$ and the reflecting boundary conditions (the bubbles in the soft zones are assumed to be open at all times) $(\partial/\partial x - 2f)P(x,y,t)|_{x=0} = (\partial/\partial y + 2f)P(x,y,t)|_{y=1} = 0$. Moreover, we impose the absorbing boundary condition $P(x, x, t) = 0$. This defines the vicious walker property [13], terminating the process when the two walkers meet.

The dynamics of any number of vicious walkers in an arbitrary but *common* potential was studied previously [14], obtaining the solution by antisymmetrization of the single-walker evolutions (Green's functions). This method, however, already fails for two walkers in two *different* potentials. Of course, equation (2) can be solved numerically as a (2+1)-dimensional partial differential equation, but it is clear that expressing the solution in terms of single-walker Green's functions would be very advantageous. It reduces the dimension of the problem by 1 and even enables us to make exact analytical statements about some aspects of the solution.

We show that for the particular problem at hand we can find the solution to equation (2) by antisymmetrization of a suitable auxiliary single-walker Green's function. This is achieved by the similarity transform of the Fokker-Planck operator into a Hermitian operator (see reference [12], chapter 5) by using $\exp(\mp f(x-x_0))\left(\frac{\partial^2}{\partial x^2} \mp 2f\frac{\partial}{\partial x}\right)\exp(\pm f(x-x_0)) = \frac{\partial^2}{\partial x^2} - f^2$ which turns both Fokker-Planck operators into the *same* Hermitian one. Together with the freedom to choose the same boundary conditions for both walkers([2]) we end up with identical differential equations for both walkers and, thus, the antisymmetrization procedure can be used for the Hermitian version of the problem. This finally results in

$$P(x,y,t) = e^{f(x-y-x_0+y_0)}\Big(p(x,t|x_0)p(y,t|y_0) - p(y,t|x_0)p(x,t|y_0)\Big), \tag{3}$$

where the auxiliary single-walker Green's function $p(x, t|x_0)$ satisfies the following 1D equation

$$\frac{\partial p(x,t|x_0)}{\partial t} = \left(\frac{\partial^2}{\partial x^2} - f^2\right)p(x,t|x_0), \quad \left(\frac{\partial}{\partial x} \mp f\right)p(x,t|x_0)\Big|_{x=1/2\mp 1/2} = 0, \tag{4}$$

([2])Each walker can reach only one boundary since the other one is blocked by the second walker. Therefore, we can choose the remaining boundary condition at will and, in particular, to be that of the second walker.

with the initial condition $p(x,0|x_0)=\delta(x-x_0)$. Equation (4) is solved by

$$p(x,z|x_0)=\left(e^{k|x-x_0|}+\kappa^2 e^{2k-k|x-x_0|}+\kappa e^{k(x+x_0)}+\kappa e^{2k-k(x+x_0)}\right)/2k\left[\kappa^2 e^{2k}-1\right], \quad (5)$$

in the Laplace domain defined by $g(z)=\int_0^\infty g(t)\exp(-zt)dt$. Here, $k\equiv\sqrt{z+f^2}$ and $\kappa=(k+f)/(k-f)$. To find the expression for the PDF $P(x,y,t)$ we need the Laplace inverse of $p(x,z)$. Note that the latter can be expanded in eigenmodes through $p(x,t)=\sum_{n=0}^\infty e^{\lambda_n t}\varphi_n(x)\varphi_n^*(x_0)$.

It is important to notice at this point that equation (4) is *not* a physically meaningful 1D Fokker-Planck equation (due to unphysical boundary conditions) and, indeed, it turns out that for $f<0$ one of the λ's, determined by the transcendent equation $(\lambda+2f^2)\sinh\sqrt{\lambda+f^2}+2f\sqrt{\lambda+f^2}\cosh\sqrt{\lambda+f^2}=0$, is actually positive, i.e., gives a solution exponentially growing in time. This type of behaviour is absent in the physical quantity $P(x,y,t)$ due to the antisymmetrization procedure (3), yet the $\lambda>0$ eigenmode plays an important role in the Kramers type asymptotics obtained for $f\ll -1$. Details are deferred to a future publication [10].

Results for the coalescence time and position. – The coalescence time PDF, $\pi(t|x_0,y_0)$, can be found using the conservation of probability

$$\int_0^t dt'\pi(t')+\int_0^1 dy\int_0^y dx P(x,y,t)=1, \quad (6)$$

where the two terms denote the probabilities that walkers have met before time t or have not, respectively. Then it follows from the Fokker-Planck equation (2) after integration by parts,

$$\pi(t)=-\int_0^1 dy\int_0^y dx\frac{\partial P(x,y,t)}{\partial t}=\int_0^1 dx\left(\frac{\partial}{\partial y}-\frac{\partial}{\partial x}\right)P(x,y,t)\Big|_{y=x}\equiv\int_0^1 dx\varrho(x,t), \quad (7)$$

where $\varrho(x,t)$ is the joint PDF for coalescence time and meeting position of the two walkers. We are interested in $\pi(t)$ and its mean value (mean coalescence time) $\tau=\int_0^\infty t\pi(t)dt$ together with the PDF of the meeting position $\rho(x)=\int_0^\infty\varrho(x,t)dt$. By standard identities of the Laplace transforms we note that $\rho(x)=\varrho(x,z=0^+)$ and $\tau=-d\pi(z)/dz|_{z=0^+}$. The related analysis becomes non-trivial as the *auxiliary* Green's function $p(x,t|x_0)$, as mentioned above, has a positive eigenvalue in the case $f<0$. Details of the calculations using the theorem of residues and high precision numerics will be presented in reference [10].

Here, we discuss the final results for the coalescence time PDF $\pi(t)$ in figure 3, and for the mean coalescence time τ and position PDF $\rho(x)$ in figure 4. In all cases the initial condition used is $x_0=0$, $y_0=1$, i.e., the walkers start out at the reflecting walls or, in terms of the original model, the bubbles start from the boundaries of the soft zones. The coalescence time probability densities graphed in figure 3 for two different values of f (or u) compare the results of all three mentioned methods (Gillespie, master equation, and the semi-analytics) and show a sharp initial raise of rather peaky structures followed by an exponential decay for long times. The short-time increase stems from the initial separation of the two forks that first have to go through several random steps before having the chance to coalesce while the long time asymptotics is fixed by the boundary conditions. The agreement between the various computational approaches is very good. The locations of the maximum probability peaks correlate well with the termination points of sample trajectories depicted in figure 2 for the same parameters.

The results for (i) $\rho(x)$ and (ii) τ calculated by the semi-analytical method are shown in figure 4. (i) The curves for the PDF $\rho(x)$ of the coalescence position exhibit a pronounced

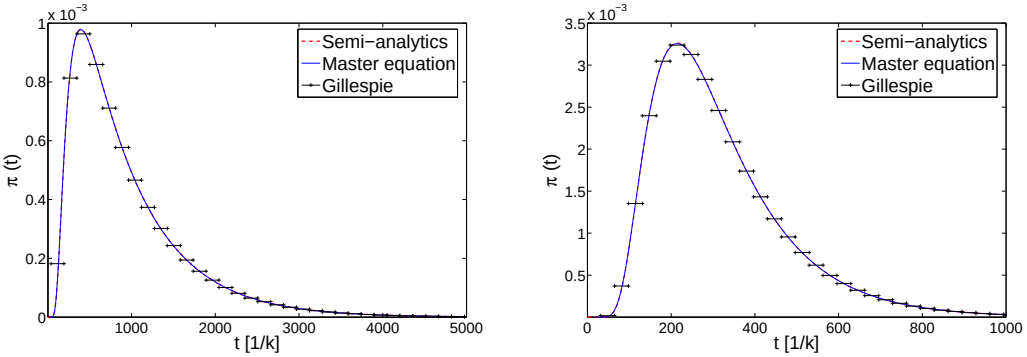

Fig. 3 – Comparison of the results for the bubble coalescence time PDF $\pi(t)$ obtained from the presented semi-analytical theory (eqs. (2)–(7)) by numerical inverse Laplace transform, the master equation, and the stochastic simulation (Gillespie), demonstrating excellent agreement. The values of the parameters are the same as in figure 2, i.e., $u = 0.98$ (left) and $u = 1.10$ (right), and $N = 50$.

crossover from a relatively sharply peaked form to an almost flat behaviour. The former occurs for large positive force f, corresponding to a strong drift toward a potential well, with negligible influence of the boundary conditions. In contrast, for large negative f, corresponding to a high barrier for coalescence, the insensitivity of $\rho(x)$ to the position x can be explained in terms of a simple Arrhenius argument: The probability of the walker to be at a position x is proportional to the Boltzmann weight, $\exp(-\beta\phi(x))$, where $\phi(x) = -\int^x F(x')dx'$ is the free energy corresponding to the force $F(x)$ (see figure 1). Then, the joint probability to have both walkers meet at the same position is given by the product $\exp(-\beta[\phi_L(x)+\phi_R(x)]) \approx$ const as the two walkers are in opposite linear potentials and the position dependence of the exponent cancels out. This simple picture necessarily breaks down close to the boundaries. (ii) The f-dependence of the mean first passage time τ crosses over from the $\tau \simeq 1/f$ behaviour typical for diffusion in a strong positive force pushing the two walkers together, to the exponential form $\tau \simeq \exp(2|f|)$ of the associated Kramers problem. The former problem was studied in reference [15] by neglecting the boundaries and switching to the relative coordinate description

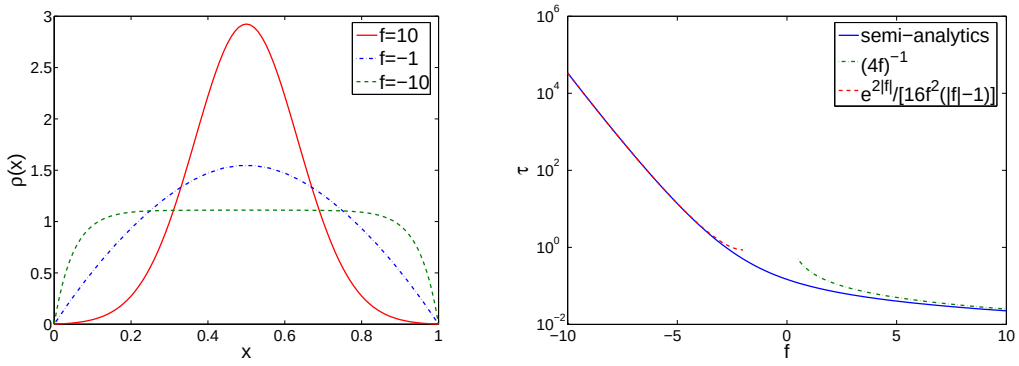

Fig. 4 – **Left:** PDF $\rho(x)$ of the coalescence position x for various values of the dimensionless force f. **Right:** Mean coalescence time τ (in dimensionless, i.e., scaled, units) as a function of f. The two analytic asymptotic behaviours for large and small f are also shown.

which enables one to find the analytic result $\tau = 1/(4f)$. For the Kramers problem ($f \ll -1$) the analytic solution for both $\rho(x) = [1 - e^{-2|f|x} - e^{-2|f|(1-x)}]|f|/(|f| - 1)$ and $\tau = e^{2|f|}/[16f^2(|f| - 1)]$ can be found rather easily [10] by the expansion into the lowest two eigenmodes of $p(x, t|x_0)$.

Conclusions. – In the present study we investigate two-bubble coalescence in a designed DNA construct consisting of two soft regions separated by a more stable barrier zone. We present a continuous semi-analytical theory yielding the coalescence time and position PDF's and show that the results agree well with the numerical results obtained with the discrete master equation and Gillespie stochastic simulation schemes. We note that for long barriers, whose sequence is arranged as random energy landscape, the propagation of the zipping forks may become subdiffusive in time [16] and the associated first passage time process is generated from the present results with the help of the subordination formalism [17]. The mathematical analysis reduces to the previously unaddressed case of two vicious walkers in *opposite* linear potentials. We demonstrate that its solution can be constructed by antisymmetrization of appropriate auxiliary one-variable densities.

* * *

The work of T. N. is a part of the research plan MSM 0021620834 financed by the Ministry of Education of the Czech Republic. R. M. acknowledges the Natural Sciences and Engineering Research Council (NSERC) of Canada, and the Canada Research Chairs programme, for support. This work was started at CPiP 2005 (Computational Problems in Physics, Helsinki, May 2005) supported by NordForsk, Nordita, and Finnish NGSMP.

REFERENCES

[1] KORNBERG, A., *DNA Synthesis* (W. H. Freeman, San Francisco, CA) 1974.
[2] POLAND, D. and SCHERAGA, H.A., *Theory of helix-coil transitions in biopolymers* (Academic Press, New York) 1970.
[3] GUÉRON, M., KOCHOYAN, M., and LEROY, J.-L., *Nature*, **328** (1987) 89.
[4] ALTAN-BONNET, G., LIBCHABER, A. and KRICHEVSKY, O., *Phys. Rev. Lett.*, **90** (2003) 138101
[5] CHOI, C. H. ET AL., *Nucleic Acids Res.*, **32** (2004) 1584.
[6] AMBJÖRNSSON, T. ET AL., *Phys. Rev. Lett.*, **97** (2006) 128105.
[7] BANIK, S. K., AMBJÖRNSSON, T., and METZLER, R., *Europhys. Lett.*, **71** (2005) 852.
[8] AMBJÖRNSSON, T., and METZLER, R., *J. Phys. Cond. Mat.*, **17** (2005) S1841.
[9] AMBJÖRNSSON, T., and METZLER, R., *Phys. Rev. E*, **72** (2005) 030901(R).
[10] PEDERSEN, J. N. ET AL., In preparation.
[11] VAN KAMPEN, N. G, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam) 1992.
[12] RISKEN, *The Fokker-Planck equation* (Springer-Verlag, Berlin) 1989.
[13] FISHER, M. E., *J. Stat. Phys.*, **34** (1984) 667.
[14] BRAY, A. J., and WINKLER, K., *J. Phys. A*, **37** (2004) 5493.
[15] HANKE, A., and METZLER, R., *J. Phys. A*, **36** (2003) L473.
[16] HWA, T. ET AL., *Proc. Natl. Acad. Sci. USA*, **100** (2003) 4411.
[17] METZLER, R., and KLAFTER, J., *J. Phys. A*, **2004** (R161) .